Protecting DOE Office of Science Resources while Maintaining an Effective Open Collaborative Science Environment

Deborah Agarwal (DAAgarwal@lbl.gov),
Dwayne Ramsey (DGRamsey@lbl.gov)

The Department of Energy (DOE) Office of Science is responsible for the operation of some of the nation's most advanced science research and development user facilities located at the national laboratories. These facilities include supercomputing centers, large-scale experiments, and the high-speed networks that connect them. DOE Office of Science researchers are also participants in experiments such as ITER, CMS, and ATLAS, which are hosted by other countries. These one-of-a-kind facilities and experiments involve thousands of scientists spread throughout the globe, including sensitive countries. In 2005, 18 of the 20 top ESnet flows were to or from an international site. Greater than 50% of the DOE Office of Science PIs and facility users are at universities and DOE’s NERSC supercomputing center has 2500 users with over 50% of these users being at universities. Many of these users rarely or never visit the DOE facility they are using.

As ESnet moves to 40 Gbps interconnections at high-end computing facilities, and the facilities themselves move to peta-scale computers, high-speed data transfers will be routinely moving peta-bytes of data to and from DOE sites. The software that supports modern open science and provides high-speed data transfers, specialized computations, distributed computational capabilities, virtual organization support, and experiment control is generally not available from commercial sources. This software has instead been developed by research and development projects to support these capabilities. One example of this largely non-commercial software is the Grid software which incorporates authentication, authorization, scheduling, data transfer, portals, etc. The Grid software forms the core of the Open Science Grid which is depended on by many science collaborations (often referred to as virtual organizations) including the Atlas and CMS experiments at the LHC.

DOE sites, as key participants in open science collaborations, need methods of participating as first-class entities and resources in the virtual organizations while protecting DOE resources from hackers. Protecting the facilities and detecting malicious attacks without adversely affecting scientific missions is particularly challenging given the high performance requirements, global user population, and diversity of custom applications and software the projects require. Securing DOE Office of Science’s high-value resources in this environment requires continual vigilance and adaptation. A recent attack on several NSF and DOE supercomputing centers is an example of the importance of protecting high-end facilities; the attack took the San Diego Supercomputer Center off the network for an entire week. Protections of high performance computers and experiments are particularly important because recovery of a compromised user facility can take weeks, during which the facility is unavailable for its mission. DOE facilities and research collaborations can ill afford to be offline for extended periods due to security incidents.

Although participating in open science while protecting DOE Office of Science resources is difficult, it can be achieved. This challenge is met by the DOE Office of Science laboratories using a variety of strategies. An example strategy for achieving both goals is the cybersecurity program at Lawrence Berkeley Laboratory (LBL). LBL (including the NERSC supercomputing center), addresses the open science cybersecurity challenge using an inside out approach to securing the systems and by teaming cybersecurity research and operations personnel. An inside out approach to security begins with the users and systems and works its way outward through the network to the site border, strengthening the defenses at all levels. Appropriate wide-area network protections such as detection and blocking of distributed denial of service attacks are provided to the site by ESnet. This approach is called defense in depth.

The LBL cybersecurity team incorporates leading-edge research, development and operational deployment in the same program. This integrated approach engages the researchers directly with the security challenges faced by the DOE open science environment and creates an innovative intellectual approach to operations. The LBL cybersecurity deployment incorporates a combination of commercial off-the-shelf technology when possible, and products of targeted research and development as needed. An example of the direct benefit of the teaming of research, development, and operations personnel is the Bro intrusion detection system which is a research and development program that is now in production use throughout LBL, NERSC and ESnet. Bro was designed specifically to address an open science environment and leverages a deep understanding of the network protocols, traffic profiles, and performance needs of an open science facility. Bro is coordinated with the site border router to allow it to put in place blocks of malicious behavior automatically. Bro also provides extensive packet capture capabilities to aid forensics. Honey Farms and the Network Telescope are also among the cybersecurity R&D projects that have benefited from this team approach and are deployed in cooperation with ESnet to aid cybersecurity operations.

A key component of LBL's cybersecurity strategy is risk analysis and return-on-investment (ROI) based decision making. Deployment of each individual component in a cybersecurity system is a balance between cost and benefit. Although best practices help inform the decision of what components to deploy, this approach, if used alone, can miss important threats not typically found in common best-practices scenarios. Calculating ROI for cybersecurity protections is still a developing science, and further development and refinement can only be accomplished by comparing real performance and cost of a protection mechanism with the expected performance. Intelligent cybersecurity deployment decisions for distributed science environments are particularly critical since the design space is almost infinite.

Cybersecurity for open science requires an evolving approach that continually keeps pace with performance requirements and the hacker environment. The next generation of cybersecurity solutions to support DOE supercomputing, user facilities, high-speed networks, laboratories, and open science will require integration of cybersecurity research, development, and operations teams to bridge the remaining gaps between the open collaborative science needs and the cybersecurity tools available today. The end goal is to provide practical solutions and tools which can be deployed operationally within research environments. Some examples of topics that need to be pursued include:

- Solutions for traffic analysis, intrusion detection, and intrusion prevention on these high-speed links. It will also create mechanisms to quickly detect and recover from intrusions on high performance computing resources.
- A framework for coordination between security components and merging information from multiple security layers to provide a more comprehensive view, enabling a more complete and accurate cybersecurity perspective.
- Tools and services to allow virtual organizations to better monitor their resources and perform incident containment.
- Vulnerability analysis and patching of custom science software.
- Tools for configuration verification and secure authenticated communication, as well as tools for aiding incident response. These are essential and often overlooked components to an overall security architecture.

A key to the success of this effort will be collaboration with networking staff, stakeholders, policy makers, and users and deployment and testing of cybersecurity infrastructure on real networks and systems. Testing over real production networks is the only means of understanding the usage scenarios and determining the utility and scalability of the approach, and is essential to the success of any security measure in this environment

The DOE national laboratories and user facilities provide an ideal environment in which to research, develop, test, and deploy an open science cybersecurity system. There is a unique level of consensus and information interchange across and within enclaves — collaboration grown from the fact that the labs and their wide-area network are all under the DOE. The DOE Office of Science labs bring together expert cybersecurity and networking operations staff and top cybersecurity researchers in an environment that allows teaming of these groups. In addition, many of the DOE science programs are based on large collaborations that span several enclaves and cross national borders. Projects such as the Open Science Grid span both DOE facilities and networks. These projects provide an ideal environment for deploying and testing interoperable cybersecurity systems within and across enclaves. The DOE Office of Science high-speed networking, supercomputing, and high-value facilities at the laboratories are extensive and need to be protected without compromising the mission of the office.

Success of a DOE Open Science Cybersecurity Research and Development Program will depend on engagement of both the research and operations communities within DOE Office of Science. The requirements and priorities for such a program needs to be gathered from operations groups within DOE Office of Science and then prioritized. One significant challenge will be long-term support and maintenance of deployed research products. A workshop that brings together key DOE Office of Science cybersecurity operations and research personnel will be convened in Washington DC on January 24-26, 2007 (http://dsd.lbl.gov/Workshops/CyberWorkshop). This workshop will further define the goals and set priorities for this cybersecurity research and development program. A program of this magnitude will require funding on the order of $25–$30 million annually to make a significant difference. DOE Office of Science through this program can take the lead and develop a model for protecting open science participation and high value resources.

Acknowlegements

This paper has developed over the course of the last two years and has benefited from the advise, comments, and edits of many people throughout the DOE and university community. The complete author and contributor list for this paper would be longer than the paper. Please direct questions and comments regarding the paper to Dr. Deborah Agarwal (DAAgarwal@lbl.gov).